Available online at www.sciencedirect.com

ScienceDirect

Procedia Social and Behavioral Sciences 9 (2010) 820–824

Procedia Social and Behavioral Sciences

WCLTA 2010

Efficiency comparison of greylisting at several SMTP servers

Tomáš Sochor [a] *

[a] *University of Ostrava, 30. dubna 22, Ostrava, Czech Republic*

Abstract

SPAM represents more than 90% of SMTP traffic in the Internet. Greylisting has become one of efficient tools for SPAM elimination since its proposing in 2003. So far only few studies focusing to greylisting efficiency were published. Greylisting implementation in Postgrey has been successfully used in the Author's university SMTP server since 2006. There were some attempts to evaluate its efficiency based on logs but due to many other aspects of SPAM elimination playing the role the direct efficiency comparison between various periods is difficult. Therefore also other SMTP servers at cooperating bodies were examined to obtain proper information about its efficiency. Two SMTP server using Postgrey with Postfix and one SMTP server using greylisting in different implementation were studied; the efficiency was estimated and mutually compared. The main result of the study is that significant efficiency decrease was not observed despite greylisting simplicity. This study extends previous findings regarding efficiency to longer period and what is especially significant, confirms results by comparing results to other sites. Also another finding demonstrating the input Postfix filtering as an efficient anti-SPAM measurement is illustrated. The main conclusion is that filtering SPAM as soon as possible, i.e. before real delivery of SPAM messages to target SMTP server, is extremely efficient anti-SPAM measure comparing other methods.

Keywords: Unsolicited e-mail messege, SPAM elimination, Greylisting, Postgrey, Postfix input filtering, SBL blocking, SPAM ratio;

1. Introduction – SPAM elimination

Unsolicited mail (hereinafter SPAM) represents the most significant problem in using electronic mail service. Supposing most e-mail users want to keep using this service in the future the application of more and more sophisticated SPAM-elimination techniques is necessary. SPAM penetration show that despite quite small share of e-mail in total Internet traffic (0.39% in 2007 – see [1]) the importance of e-mail is very high especially in businesses and governmental agencies. Various sources show that huge percentage of e-mail traffic is represented by SPAM (e.g. [2] estimated SPAM share in 2005 to more than two third) and the SPAM share seems to be still growing as also our data show. The data from our measurements show that the SPAM ratio recently exceeds 90% (see Fig. 1 – upper longer line). An additional note should be added: the diagram in the Fig. 1 shows only the part of SPAM messages (however significant majority), namely the messages that were identified as SPAM. Despite the fact that multilevel SPAM detection techniques as described below are very efficient there is still certain amount of SPAM messages undetected. This is usually necessary especially due to the fact that it is known that too strict adjustment of SPAM detecting mechanisms (similar to antivirus scanners) increases the risk of "false positive"

* Corresponding author. Tel.: +420-59-7092129; fax: +420-59-7092264.
E-mail address: tomas.sochor@osu.cz.

1877-0428
doi:10.1016/j.sbspro.2010.12.241

detections, i.e. situations when “legal” message is detected as SPAM. Such situation is considered as much more adverse than passing several SPAM messages every day to the users mailbox undetected.

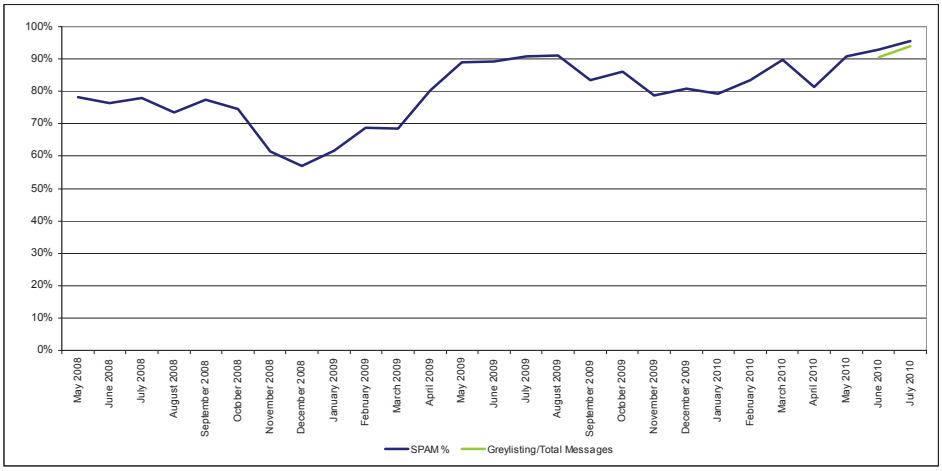

Fig. 1. Evolution of SPAM ratio for recent period (incl. greylisting share in last two months - lighter line) – SMTP server Zebra

1.1. SPAM elimination approaches

There are numerous approaches to the SPAM elimination issue. For the purpose of this text the SPAM elimination methods will be classified into two categories:

- methods detecting SPAM having been delivered to the target SMTP server, and
- methods detecting SPAM in advance of the message delivery.

The first approach is usually applied in traditional SPAM filters. The obvious advantage of the first approach is in the fact that the complete message is available and complete message code could be analyzed. This factor often results in possible more precise SPAM detection. Also time available for processing is not so critical in this case. On the other hand the fact that message is delivered first and then could be identified as SPAM resulting either in its deleting or at least its delivery in a form different from “legal” messages to the user's mailbox could result in loss of the message that originator consider to be delivered.

The second approach suffers from limited data about the message being delivered (usually just the data taken from the SMTP dialogue as described in [3]). Such data obviously do not allow application of content-based filtering methods. The significant advantages of this approach are both increase of e-mail service reliability (in the sense that this approach does not produce “false delivered” messages) and considerable savings of SMTP server resources (due to the fact that SPAM messages need not be delivered and saved). This approach can be represented by greylisting that will be described later in details. Both approaches are often used in combination. The combination used in one of studied sites is shown in the Fig. 2.

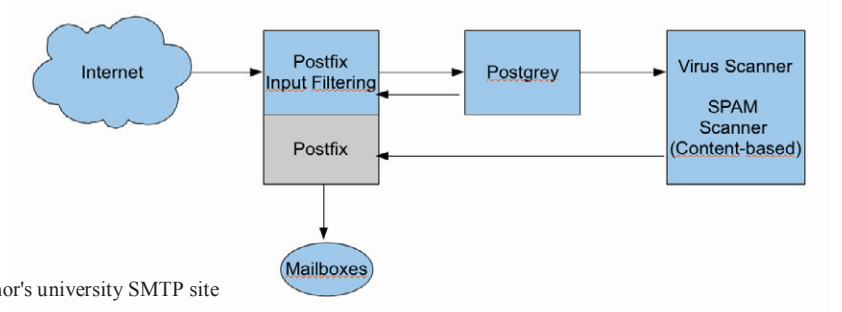

Fig. 2. Layout of anti-SPAM system at the author's university SMTP site

1.1. Greylisting

The idea of greylisting is very simple – temporary blocking of messages from new sources that is expected to be no problem for legal SMTP servers but could be difficult to overcome for SPAM producers. It was proposed in 2003 (see [4]). Their behavior was studied in our previous works (e.g. [5]) but all previous studies were limited to the single site. This study tries to compare results from three different SMTP servers used in various situations.

2. Methods

The data from SMTP servers (including greylisting implementation if external) are continuously logged and such logs are processed (after necessary anonymization). Two of studied SMTP servers use Postfix SMTP server (in description of figures the SMTP server of Author´s University identified as "OU", the other identified as "UKF") with Postgrey as policy server in a way similar to the Fig. 2. The remaining site uses MS Exchange with commercial antivirus and anti-SPAM solution by GFI software where greylisting was implemented in the first half of 2010 (in description of figures identified as "Zebra").

The processing of logs was dependent on the amount of data available in logs but in case of Postfix and Postgrey logs were stored into a database to allow association of individual events associated with single e-mail message. From GFI software standard statistic reports were used.

3. Results

The SMTP server of the author's university has been using greylisting since the end of 2006. As the diagram in the Fig. 3 shows in 2007 greylisting itself (supported just by simple input Postfix filtering) was able to eliminate as much as 85 – 90 % of SPAM. From Jan 2008 the ratio started decreasing (dotted line in Fig. 3). But this drop is not result of decreasing greylisting efficiency. As one can see e.g. in the Fig. 1, greylisting is able to filter out more than 90% of SPAM even in 2010.

The decrease of the dotted line in the diagram in the Fig. 3 at the beginning of 2008 is due to implementation of Intrusion Preventing System (IPS TippingPoint by 3Com). This IPS operating at the outer edge of the network blocks certain attempts to connect to computers in our network including certain part of attempts to deliver SPAM to the SMTP server. Because of the nature of the IPS system it would be very complicated and system-resources consuming to get detailed logs of such blocked attempts (and their analysis could be very complicated too) there is no detailed analysis of the IPS effect.

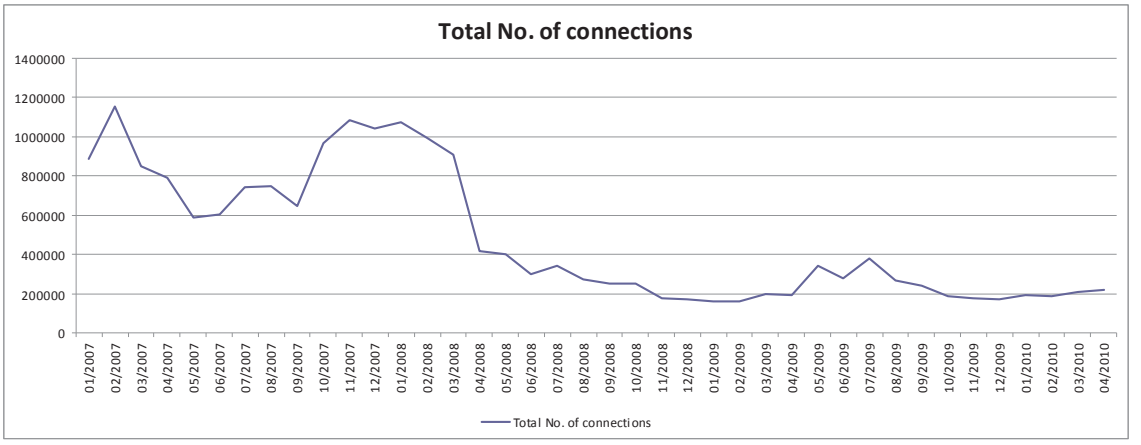

Fig. 3. Percentage of messages blocked to pass through Postgrey related to the total amount of attempts to deliver a message registered by Postfix – SMTP server OU (dotted line – original data, solid line – ratio recalculated against the average number of incoming messages in 2007.

Moreover the IPS is does not behave like static filter but it has some self-learning abilities so that its efficiency in blocking unwanted traffic is very likely to increase. This increase reflects in gradual (though irregular) increase of IPS blocking efficiency. This is also documented by the diagram in the Fig. 4 where total amount of delivery attempts to postfix gradually grow up to the top in Jan. 2008 and then the values start to decrease. Therefore also solid line has been added in the Fig. 3 expressing the refusal ratio recalculated again the average of incoming messages in 2007. Thus the influence of decrease of incoming messages number was eliminated and it is clear that the efficiency (refusal ratio) remains above 90 %.

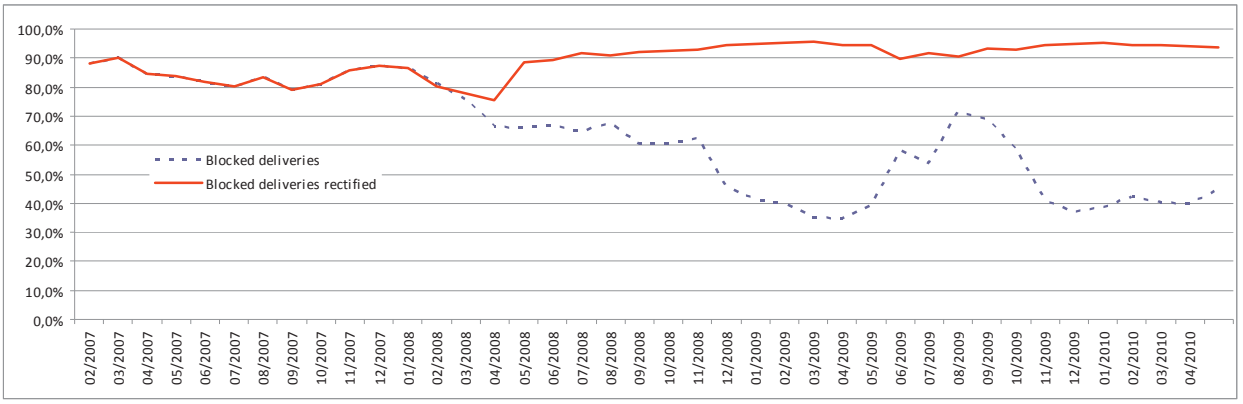

Fig. 4. Number of message delivery attempts on monthly basis 2007 – 2010 – SMTP server OU.

In the Fig. 3 there is another stepwise decrease from Oct. 2009. It is due to the fact that starting this month the Realtime Blackhole List blocking was implemented in our Postfix. For this blocking the SBL database of Spamhaus.org is used (see [6]). The efficiency of this blocking is demonstrated in the diagram in Fig. 5.

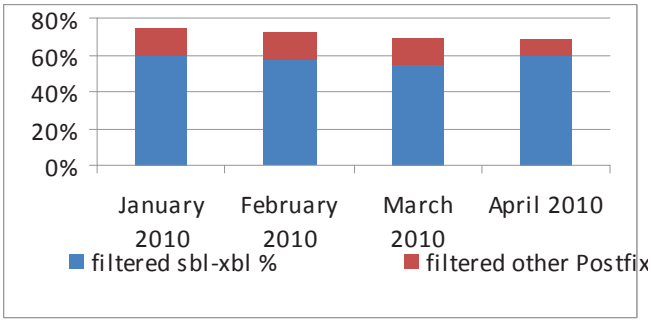

Fig. 5. Percentages of blocked messages on Postfix input by RBL and by other Postfix filters in 2010 – SMTP server OU.

The data shown in the diagrams above could seem not to prove the efficiency of greylisting to be stable. Therefore another diagram is available showing the number of messages delivered using AWL. The AWL is an acronym for Automatic White List and it is a list of IP addresses from which several messages (adjustable value, usually set to 5 and this value is used in our university PostGrey) were successfully delivered during a month. The number of messages passing by Postgrey based AWL reflects the number of “trusted” sources. If the number increases it could be a sign of Postgrey efficiency decay. But as shown in the Fig. 6 this value remains relatively stable (except several fluctuations but the average value does not tend to increase or decrease) for the whole period.

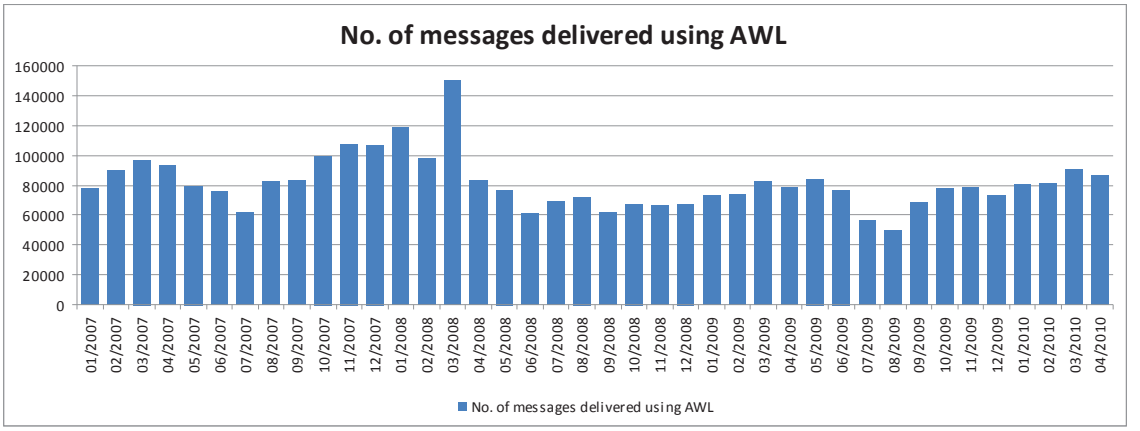

Fig. 6. Monthly number of messages delivered immediately based on the originators IP address in AWL – SMTP server OU.

Unfortunately due to various technical and other reasons the amount of data (their detailedness and/or time range) is significantly smaller than for OU. Due to the fact the comparison is limited to comparing just few figures.

The most important parameter of greylisting is efficiency. As show above the initial efficiency in OU was almost 90% and the subsequent decay is due to other factors so that the greylisting efficiency at OU still remains between 80 and 90%. As data from Zebra shows (see Fig. 1 lower shorter line) the greylisting efficiency is 90.4% for June 2010 and 93.9% for July 2010. In UKF server the ratio between number of allowed messages and number of connections is 4.99% resulting in efficiency exceeding 95% (data for April 2010).

4. Conclusions

The figures from OU above confirm that greylisting is still very efficient anti-SPAM tool despite that fact that other methods assume significant part of filtering. The comparison of OU data to other analyzed SMTP sites show is remaining sites greylisting remains to be very efficient while it era the main load of filtering there.

Acknowledgements

The author greatly appreciates the courtesy of the Zebra Systems company (Ostrava, Czech Republic) as well as the University of Nitra (Slovakia) which provided statistics from their anti-SPAM systems for purpose of this study.

References

1. Ipoque, Internet Study 2007. [online]. 2008 [quot. 2010-08-10].. Available at WWW: <http://www.ipoque.com/resources/internet-studies/internet-study-2007>

2. J. Goodman, D. Heckerman, R. Rounthwaite,. Stopping Spam. Scientific American 2005; 292(4): 42–49.

3. J. Klensin, Simple Mail Transfer Protocol. IETF [online]. Oct 2008 [quot. 2010-06-09]. Available at www: <http://tools.ietf.org/html/rfc5321>

4. E. Harris, The next step in the spam control war: greylisting. [online] 2003. [quot. 2010-06-09]. Available at www: <http://projects.puremagic.com/greylisting/whitepaper.html>

5. T. Sochor. Greylisting method analysis in real SMTP server environment : Case-study. In T. Sobh (Ed.), Innovations and Advances in Computer Sciences and Engineering. Dordrecht : Springer, 2010. pp. 423-427. ISBN 978-90-481-3657-5.

6. The SPAMHAUS Project [online]. 2010 [quot. 2010-08-13]. Available at WWW: <http://www.spamhaus.org/sbl/>..